Does Generative AI Understand Heat? Why Physics Should Matter in the AI Era.

Figure 1: Editing an image of the Stanford Bunny using GenAI. While relighting results in the visible spectrum are striking, translating to thermal domain produces a visually rich image that is physically inaccurate. Moreover, the actual thermal image reveals residual heat from my touch, information absent in the visible image in the first place.

Generative AI (GenAI) has advanced so rapidly that it now impresses even the skeptics. When I asked Gemini to relight my tabletop scene (Fig. 1), the realism of its output was striking. Then I asked it to generate what a thermal camera would capture. It produced a visually rich image, but one that diverged from the actual thermal image of the scene. First, **its thermal appearance was *unrealistic*** - the high frequency textures on the surface should have been smoothed out due to heat conduction. Second, **it was *inconsistent* with the input image** - the model seems to hallucinate an internal heat source instead of an object in thermal equilibrium. Third, **it was *inconsistent* with the actual thermal image** which shows residual heat from when I touched it. The latter is not a limitation of the model; rather, it illustrates novel information available in the thermal modality but absent in the visible modality.

But why care about thermal? Because heat is as fundamental as light. It is arguably more pervasive, yet invisible to the human eye, both literally and metaphorically. Current models can **mimic visible *appearance*** through memorization, but heat is inseparable from ***function* and *history***. A leaf after photosynthesis or a tumor after metabolism dissipates heat revealing *function* even when its visual appearance remains unchanged. Similarly, a strolling cat or a cold draft triggers heat transport that persists long after its *cause* has left the field of view, thus retaining *history*. Thus, ***thermal appearance* gives you a direct cue** about *function* and *history* for which current AI models using *visible appearance* must rely on indirect cues like dirty footprints or a waving curtain, if at all they exist. So before we jump to make AI understand heat, it is instructive we understand what the physics of heat and thermal appearance reveals about the scene, just as early computer vision researchers explored for the visible appearance more than 50 years ago.

Visible light is only a thin slice of the electromagnetic spectrum — and even all light combined is just one form of energy. My work explores heat, another fundamental form of energy, as an untapped carrier of *visual information*. I have explored physics-based thermal vision along three key axes:

1. **Theory:** *Absorbed light* carries away information that never reaches the camera, but my work [14] shows that it can be recovered from heat to solve fundamental *ill-posed* problems in computer vision. Building on this, I have modeled the three interconnected transport phenomena — visible light [18], heat [11, 12], and thermal light [10] — laying the foundation for a new class of *functional imaging* systems I will build.

2. **Measurements:** Having built diverse datasets [7, 3] that push perception boundaries, I have a deep understanding of the physics of novel sensors, especially thermal cameras. My work [15] revealed how *thermal inertia* causes motion blur in microbolometers. I've modeled them from pixel-level physics to readout pipelines [2]. I use noisy, low-cost sensors to study even subtle heat transients that permeate our world.

3. **Inference:** I translate theoretical models into practical algorithms using diverse techniques from classical sparsity priors [15] and regularization losses [11, 10, 12] to neural implicit representations [1, 6, 2] and structural priors of neural networks [18]. I view inference as a continuous optimization process and believe the temporal dimension is underutilized due to a hardware-software mismatch.

As an expert in thermal imaging, I have served as a **qualified external reviewer for NASA's PSTAR (Mapping) program** and will be teaching a **course on Light and Heat at SIGGRAPH Asia**, a premier graphics conference. I look forward to establishing an interdisciplinary research group, **MARVILS**, dedicated to advancing intelligence under the unified view of light, heat, and, more broadly, energetics. Our initial thrusts will build core capabilities across three pillars: theory **(volumetric light and heat transport)**, measurements **(alternative visible and thermal sensors)**, and inference **(foundation models for light and heat)**. These advancements would benefit diverse applications including robotics, healthcare, food security and disaster mitigation. Beyond these tangible problems, **heat has a surprising connection to causality** that aligns my work with both sponsorship priorities and the aspirations of the larger student body.

Could heat enable *causal reasoning*?

My research opens a new tool, heat, to help with one of AI's hardest problems: **causality**. Heat is not a nuisance byproduct — it carries *irreversible traces of cause and effect*. Physicists have argued that causation is rooted in thermodynamics [17] which gives us the arrow of time [16]. Consider a light source heating up an object → its thermal radiation increases → and temperature persists even after the light is turned off. This persistence makes temperature a useful latent variable that binds past events to present appearance. We could time reverse the visible video and notice no change. But if we time reverse the thermal video, it would show the object heating up before the light is turned on and cooling down afterwards, which is simply impossible.

White House's priorities [4] call for AI that is interpretable, controllable, and grounded in the sciences — underscoring the relevance of exploring the heat-causality link. So far, research in explainable AI draws inspiration from thermodynamics [9] and Physics-Informed Neural Networks [13] embed physical laws into network architectures or loss functions for specialized problems. I propose to take a complementary, pragmatic approach: understanding heat transport from thermal videos of *everyday scenes*. This connects complex science (Navier-Stokes equations for fluid dynamics) to tangible applications, say a smart thermostat powered by a thermal camera. Just as GenAI excels at relighting without explicitly using Radiative Transfer Equation, billions of thermal videos of air vents might "teach" AI to internalize fluid dynamics. More broadly, I believe AI must move beyond mimicking appearance to analyzing **energy transport**. My group, MARVILS, will develop **Vision-By-Energetics (ViBE)**, a conceptual framework I envision as a compiler for Causal Intelligence.

[Theory] We are in the Retinex and Shape-from-Shading Phase for Thermal Vision

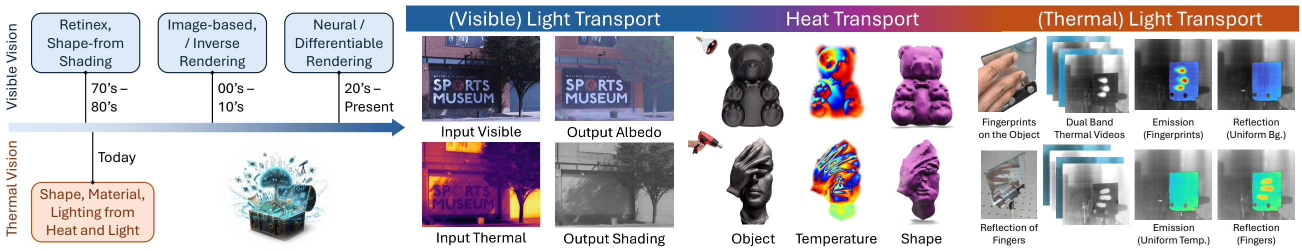

Figure 2: (Left) Timeline comparing physics-based vision research across visible and thermal spectra. (Right) Collage of results — Intrinsic Image Decomposition [18], Shape from Heat [11], Reflection-Emission Separation in thermal [10] — leveraging the joint theory of light and heat introduced in [14].

Imagine doing computer vision in the 1970s. There were no deep networks, no GPUs, and not even digital cameras. Vision was a puzzle, and the tools were physics and mathematics. Foundational ideas like **Retinex** [8] and **Shape-from-Shading** [5] defined what was possible — and what was not — when recovering shape, material, and lighting from a single photograph, establishing that *"Intrinsic image decomposition (IID) is ill-posed"*. Recently, we've revisited these foundational questions with a new tool: **a thermal camera**.

We pursued a simple question: *What is the physical relationship between visible and thermal pixel intensities?* We have shown that: i) IID becomes *well-posed* when absorbed light intensity is known [14, 18], ii) Heat conduction reveals shape even for dark, transparent or translucent objects [11], and iii) thermal videos can be decomposed into reflected and emitted radiance [10]. We are currently building collaborations with colleagues at CMU, UW-Madison, and NVIDIA on building physically accurate light-heat simulators.

To the uninitiated, thermal cameras might evoke blurry, low resolution thermograms. But today thermal cameras are commodity sensors, available as phone/robot/drone accessories. While thermal images have long been used in vision and robotics, they are often treated as grayscale intensity maps — without modeling their underlying physics. Our work opens the door to translate decades of progress in computer vision onto the thermal spectrum — essentially a treasure trove of ideas waiting to be executed. To facilitate these algorithmic advancements, my group will also build the next generation of thermal imaging systems.

[Measurements] Thermal Imaging needs its iPhone moment

Figure 3: (Left) Timeline of camera technology evolution across visible and thermal spectra. (Middle) Motion blur in uncooled microbolometers — caused by thermal inertia — can be mitigated using per-pixel non-blind temporal deconvolution for arbitrary camera motion, scene motion, and depth [15]. (Right) Iterations of a multi-sensor rig we developed for autonomous driving research in adverse weather conditions [3].

Thermal cameras (uncooled microbolometers) were invented alongside digital cameras. While visible cameras achieved explosive adoption, thermal cameras — despite becoming compact, rugged commodity sensors — await their **iPhone moment.** A key reason is that microbolometers operate on fundamentally different physics with no controllable exposure time, and uncontrollable motion blur.

These devices defy our intuition: when a visible camera records a sharp image at 30 fps, why does the corresponding thermal image look blurred — even at 200 fps? It is because each of their pixels is a temperature-dependent resistor that is always exposed to the scene — heating up and cooling down in response to scene change. It is essentially a low-pass filter acting on incident radiance. This *thermal inertia* at each pixel manifests as motion blur. While computer vision models blur as a spatial convolution with an unknown kernel that depends on camera motion, scene motion and depth, we showed [15] that **motion deblurring in thermal cameras is a per-pixel temporal deconvolution problem** with a fixed kernel calibrated *a priori*.

Throughout my doctoral studies, I have worked extensively with novel sensors, collected data in adverse weather conditions, and collaborated closely with researchers at the Ford Motor Company as well as FLIR (now Teledyne), the largest manufacturer of thermal cameras. While visible cameras boast tens of megapixels, 0.3MP is already considered high for thermal imaging. I don't believe chasing faster frame rates or higher resolution is the right objective here. Instead, the time is right for **thermal imaging to** skip the smartphone phase and **leapfrog directly to the AI-powered camera era.** MARVILS' unique expertise in both camera physics and the physics of light and heat transport positions us to best uncover latent scene information encoded in its thermal appearance — transforming what was once a fancy temperature sensor into a rich source of visual information.

[Inference] Inferring Hidden Scene Functions from Thermal Appearance

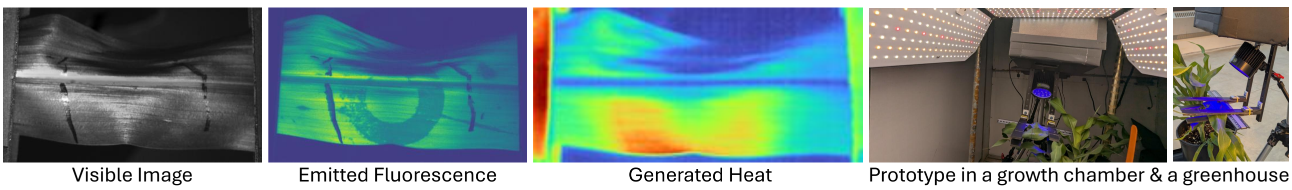

Figure 4: From left to right: a) Visible image of a leaf is dominated by surface appearance. b) Chlorophyll fluorescence reveals regions with potential photoinhibition due to a leaf clip. c) Intensity of generated heat shows complementary information. (d & e) Our multi-sensor prototype being used in field trials at Iowa State University.

Heat transport is intimately tied to an object's function — even when light transport, especially in the visible, shows no sign of activity. This is particularly striking in the case of **photosynthesis**. As part of my work with the **AI Institute for Resilient Agriculture (AIIRA)**, I have been developing a prototype that uses **controlled lighting** and a **pair of fluorescence and thermal cameras** to study energy transduction on plant leaves. Our work has drawn strong interest from plant scientists at the **University of Nebraska - Lincoln**, **Iowa State University** and **University of Arizona**. This work introduces novel theory to estimate photosynthesis and is nearing experimental validation on this one-of-a-kind problem. During the annual review, I had the opportunity to **present this work to program leaders from USDA-NIFA and NSF AI Institutes** — an educational experience that offered candid insights into the mechanics behind funding decisions.

Just like photosynthesis, metabolism also produces thermal signatures — opening exciting possibilities for applications in **cancer research**. Even among inanimate objects, **thermal appearance can reveal phenomena that are otherwise invisible**. While light cannot penetrate an opaque surface, heat can – allowing us to see what someone writes on the board from the back side. Similarly, thermal imaging exposes internal support structures in opaque 3D printed objects due to spatial differences in their effective density. My ongoing work with physics-based denoising of low-cost thermal sensors is able to extract even subtle structures with unprecedented detail. These examples underscore a profound shift: **thermal appearance is** not just about visual information — it is **a window into hidden functions and structures**. At MARVILS, we will continue to expand this frontier.

Mission: Vision-By-Energetics (ViBE) as the Compiler for Causal Intelligence

I am eager to help shape a future where the abstract reasoning power of modern AI is firmly grounded in the tangible physical quanta of energetics. Beyond the short-term projects I identified in each of the previous sections, I sketch out my long-term vision below. For a multi-modal, agentic future, where agents, not humans, interpret the raw data, our research will focus on three complementary axes:

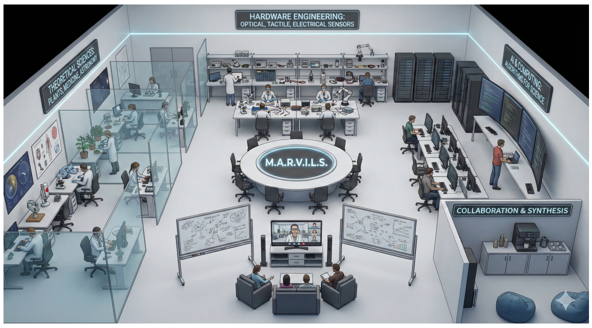

Figure 5: A render of MARVILS, my future lab. I envision it as a dynamic hub for incubating ideas and experiences, where students enter with curiosity and passion and emerge with clarity, confidence, and purpose.

A. Blending Electromagnetics (EM) and Thermodynamics (TD): Research across the sciences — plants, medicine, astronomy — routinely leverages both light (EM) and heat (TD). The theoretical wing of MARVILS will work with diverse domain experts to align our modeling capabilities with their scientific needs. This positions us to bridge disciplinary silos, paving the way to *rebase* AI onto a unified energetics framework **where light and heat are interdependent in a shared causal system**.

B. Move towards Adaptive Multiphysics Sensing: Modern cameras chase more pixels and higher throughput, but when GenAI can recover an image from jut 1% of pixels, the real goal is capturing novel information. The hardware wing of MARVILS will work with experts in optics and semi-conductors to co-design heterogeneous sensors that sample the plenoptic function (and its derivatives). By bridging their fabrication capabilities with our modeling capabilities, we create a lyer of abstraction between sensor physics and downstream vision and reasoning algorithms. This will enable perception systems **that sense adaptively — capturing what matters, when it matters, and how it matters.**

C. Elevating AI to reason about Function and Causality: As we embed into diverse scientific domains and build novel hardware, the inference wing of MARVILS will develop the *ViBE compiler* that takes *energetic signals* and produces *symbolic tokens*. To translate the success of LLMs from the symbolic world to the physical world, we must **cover all of nature's languages**, i.e. all forms of energy. This would enable AI to reason not just about appearance, but also about function and causality.

Back in 2020, I ended my 3-minute **Robotics Science and Systems (RSS) Pioneer**[1] pitch with the question: *"What if robots evolved their own visual system?"* This quest led me to purse postdoctoral training in computational imaging and taught me what a camera is from first principles. Today, I ask a broader question: *"What if superintelligence designs its own sensing suite?"* My closest answer is to imagine the laws of physics embodied as J.A.R.V.I.S. from Marvel comics.

Come, let's build J.A.R.V.I.S. together!

[1] Link to my RSS Pioneers Pitch in 2020

References

[1] A. Carlson, **Mani**. Ramanagopal, N. Tseng, M. Johnson-Roberson, R. Vasudevan, and K. A. Skinner. Cloner: Camera-lidar fusion for occupancy grid-aided neural representations. *IEEE Robotics and Automation Letters*, 8(5):2812–2819, 2023.

[2] S. Carmichael, M. Bhat, **Mani**. Ramanagopal, A. Buchan, R. Vasudevan, and K. A. Skinner. Trnerf: Restoring blurry, rolling shutter, and noisy thermal images with neural radiance fields. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 7980–7990. IEEE, 2025.

[3] S. Carmichael, A. Buchan, **Mani**. Ramanagopal, R. Ravi, R. Vasudevan, and K. A. Skinner. Dataset and benchmark: Novel sensors for autonomous vehicle perception. *The International Journal of Robotics Research*, 2024. 02783649241273554.

[4] Executive Office of the President, Office of Science and Technology Policy and Office of Management and Budget. Fiscal year (fy) 2027 administration research and development budget priorities and cross-cutting actions, September 2025. OMB Memorandum M-25-34, NSTM-2.

[5] B. K. P. Horn. Obtaining shape from shading information. In P. H. Winston, editor, *The Psychology of Computer Vision*, pages 115–155. McGraw-Hill, 1975.

[6] S. Isaacson, P.-C. Kung, **Mani**. Ramanagopal, R. Vasudevan, and K. A. Skinner. Loner: Lidar-only neural representations for real-time slam. *IEEE Robotics and Automation Letters*, 8(12):8042–8049, 2023.

[7] W. Kim, **Mani**. Ramanagopal, C. Barto, M.-Y. Yu, K. Rosaen, N. Goumas, R. Vasudevan, and M. Johnson-Roberson. Pedx: Benchmark dataset for metric 3-d pose estimation of pedestrians in complex urban intersections. *IEEE Robotics and Automation Letters*, 4(2):1940–1947, 2019.

[8] E. H. Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[9] S. Mehdi and P. Tiwary. Thermodynamics-inspired explanations of artificial intelligence. *Nature Communications*, 15:51970, 2024.

[10] S. Narayanan, **Mani**. Ramanagopal, and S. G. Narasimhan. Dual band video thermography near ambient conditions. *arXiv*, 2025.

[11] S. Narayanan, **Mani**. Ramanagopal, M. Sheinin, A. C. Sankaranarayanan, and S. G. Narasimhan. Shape from heat conduction. In *European Conference on Computer Vision (ECCV)*, pages 426–444. Springer, 2024.

[12] A. Oharazawa, S. Narayanan, **Mani**. Ramanagopal, and S. G. Narasimhan. Resolving shape ambiguities using heat conduction and shading. In *2025 IEEE International Conference on Computational Photography (ICCP)*, pages 1–11, Los Alamitos, CA, USA, July 2025. IEEE Computer Society.

[13] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[14] **Mani**. Ramanagopal, S. Narayanan, A. C. Sankaranarayanan, and S. G. Narasimhan. A theory of joint light and heat transport for lambertian scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11924–11933, 2024.

[15] **Mani**. Ramanagopal, Z. Zhang, R. Vasudevan, and M. Johnson-Roberson. Pixel-wise motion deblurring of thermal videos. In *Proceedings of Robotics: Science and Systems (RSS)*, Corvallis, Oregon, USA, July 2020.

[16] H. Reichenbach. *The Direction of Time*, volume 65. University of California Press, 1991.

[17] C. Rovelli. How causation is rooted in thermodynamics. *arXiv*, 2022.

[18] Z. L. Yuan, **Mani**. Ramanagopal, A. C. Sankaranarayanan, and S. G. Narasimhan. Ordinality of visible-thermal image intensities for intrinsic image decomposition. *arXiv*, 2025.